

## CLAIMS

What is claimed is:

5     1 <sup>sub</sup> A program storage device readable by a machine, tangibly  
embodying a program of instructions executable by the  
machine to perform method steps for constructing  
segmentation-based models that satisfy constraints on the  
statistical properties of the segments, the method steps  
10     comprising:

- 15     (1) presenting a collection of training data records  
comprising examples of input values that are  
available to the model together with the  
corresponding desired output value(s) that the  
model is intended to predict;

and

- 20     (2) generating on the basis of the training data a  
plurality of segment models, that together comprise  
an overall model, wherein each segment model is  
associated with a specific segment of the training  
data, the step of generating comprising performing  
25     optimization steps comprising:

- 30     a) generating alternate training data segments  
and associated segment models;
- b) evaluating at least one generated segment  
to determine whether it satisfies at least

one statistical constraint comprising a test whose outcome is not equivalent to a comparison between, on the one hand, the number of training records of at least one species of training records belonging to the segment and, on the other hand, a numerical quantity that may depend on the combination of species of training records being considered but that is otherwise constant for all generated segments that are evaluated;

and

c) selecting a final plurality of segment models and associated segments from among the alternates evaluated that have satisfactory evaluations.

2. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for constructing segmentation-based models that satisfy constraints on the statistical properties of the segments, the method steps comprising:

(1) presenting a collection of training data records comprising examples of input values that are available to the model together with the

corresponding desired output value(s) that the model is intended to predict;

and

(2) generating on the basis of the training data a plurality of segment models, that together comprise an overall model, wherein each segment model is associated with a specific segment of the training data, the step of generating comprising performing optimization steps comprising:

a) generating alternate training data segments and associated segment models using statistical constraints to guide the construction of the data segments in a closed-loop fashion so as to ensure that the resulting data segments satisfy the statistical constraints;

and

b) selecting a final plurality of segment models and associated segments from among the alternates generated.

3. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for constructing segmentation-based models that satisfy constraints on the

statistical properties of the segments, the method steps comprising:

5 (1) presenting a collection of training data records comprising examples of input values that are available to the model together with the corresponding desired output value(s) that the model is intended to predict;

10 (2) generating on the basis of the training data a plurality of segment models, that together comprise an overall model, wherein each segment model is associated with a specific segment of the training data, the step of generating comprising:

15 a) generating alternate pluralities of data segments and associated segment models;

20 and

b) adjusting the alternate pluralities so that the resulting data segments satisfy the statistical constraints.

25 4. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform method steps for constructing segmentation-based models of insurance risks, the method  
30 steps comprising:

(1) presenting a collection of training data comprising examples of historical policy and claims data;

and

(2) generating on the basis of the training data a plurality of segment models, that together comprise an overall model, wherein each segment model is associated with a specific segment of the training data, the step of generating comprising performing optimization steps comprising:

a) generating alternate training data segments and associated segment models;

b) evaluating the generated segment models using numerical criteria derived from statistical models used by actuaries to model insurance risks,

and

c) selecting a final plurality of segment models and associated segments from among the alternates generated so as to optimize aggregate numerical criteria for the plurality.

add  
G2